

中華民國經濟部智慧財產局

INTELLECTUAL PROPERTY OFFICE
MINISTRY OF ECONOMIC AFFAIRS
REPUBLIC OF CHINA

茲證明所附文件，係本局存檔中原申請案的副本，正確無訛，
其申請資料如下：

This is to certify that annexed is a true copy from the records of this
office of the application as originally filed which is identified hereun

申請日：西元 2002 年 05 月 29 日
Application Date

申請案號：091111432
Application No.

申請人：艾爾科技股份有限公司 Technology Center 2600
Applicant(s)

RECEIVED

OCT 07 2002

CERTIFIED COPY OF
PRIORITY DOCUMENT

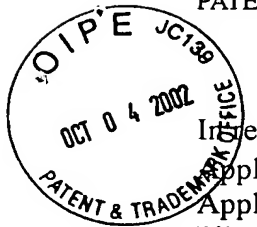
局長

Director General

蔡練生

發文日期：西元 2002 年 9 月 10 日
Issue Date

發文字號：09111017319
Serial No.



PATENT

Customer No. 31561
Attorney Docket No. 08990-08991

PIN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the application of
Applicant : Yi-Jing Lin
Application No. : 10/064,616
Filed : 2002/7/31
For : USER INTERFACE, SYSTEM, AND METHOD FOR
AUTOMATICALLY LABELLING PHONIC SYMBOLS TO
SPEECH SIGNALS FOR CORRECTING PRONUNCIATION
Examiner :

ASSISTANT COMMISSIONER FOR PATENTS

Washington, D.C. 20231

RECEIVED

OCT 07 2002

Technology Center 2600

Dear Sirs:

Transmitted herewith is a certified copy of Taiwan Application No.: 91111432,
filed on: 2002/5/29.

A return prepaid postcard is also included herewith.

Respectfully Submitted,
JIANQ CHYUN Intellectual Property Office

Dated: *October 2, 2002*

By: *Belinda Lee*
Belinda Lee
Registration No.: 46,863

Please send future correspondence to:
7F.-1, No. 100, Roosevelt Rd.,
Sec. 2, Taipei 100, Taiwan, R.O.C.
Tel: 886-2-2369 2800
Fax: 886-2-2369 7233 / 886-2-2369 7234

申請日期	
案 號	91111432
類 別	

A4
C4

(以上各欄由本局填註)

發 明 專 利 說 明 書

一、發明 名稱	中 文	自動標示音標以矯正發音之使用者介面、系統及方法
	英 文	INTERFACE OF AUTOMATICALLY LABELING PHONIC SYMBOLS FOR CORRECTING USER'S PRONUNCIATION, AND SYSTEMS AND METHODS
二、發明 創作人	姓 名	林宜敬 Yi-Jing Lin
	國 籍	中華民國
三、申請人	住、居所	台北市建國北路一段 124 巷 3 號 5 樓
	姓 名 (名稱)	艾爾科技股份有限公司 L LABS Corporation
	國 籍	中華民國
	住、居所 (事務所)	台北市松江路 82 號 8 樓 -1
	代 表 人 姓 名	林宜敬 Yi-Jing Lin

裝

訂

線

四、中文發明摘要(發明之名稱：自動標示音標以矯正發音之使用者
介面、系統及方法)

一種自動標示音標以矯正發音之使用者介面、系統及方法，係利用圖形介面比較並顯示語言學習者與語言教學者在發音上的差異，以幫助語言學習者矯正發音的電腦系統，包含其使用者介面、系統及方法。當使用者提供一個文句字串並輸入其相對的聲音訊號之後，本發明會在輸入聲音訊號的各個區段標示出其對應的音標。接著以該些區段之對應音標，本系統可以比對出教學者聲音訊號與學習者聲音訊號之間的差異，如各個音節的發音、音高、強度、長短的差異等，並據此進行評分或提出改善建議。

英文發明摘要(發明之名稱：

INTERFACE OF AUTOMATICALLY
LABELING PHONIC SYMBOLS FOR CORRECTING USER'S
PRONUNCIATION, AND SYSTEMS AND METHODS

An interface of automatically labeling phonic symbols for correcting user's pronunciation, and systems and methods, which are implemented by a computer system. The computer system uses a graphic interface to automatically compare and showing difference between the learner's pronunciation and the demonstrator's pronunciation, in order to help the learner correcting his pronunciation. When a string of phonic symbols are

四、中文發明摘要 (發明之名稱：)

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

訂

線

英文發明摘要 (發明之名稱：)

input from the user, frames of the input string of phonic symbols are labeled by corresponding phonic labels. By comparing the corresponding labels of the frames of the phonic symbols, the system can obtain difference between the phonic symbols of the learner and phonic symbols of the demonstrator originally stored in the computer system, in order to correct the required speed, pitch, energy and articulation of each vocabulary of the learner.

五、發明說明 (/)

發明領域

本發明是有關於一種矯正發音系統之使用者介面、製造及使用方法。其特點在於能快速而正確的標示出一個聲音訊號的各個音節的音標，並據此比較出語言教學者與語言學習者在發音上的差異，進而提出改善建議。

發明背景

當人們學習外語的時候，不外乎是學習該語言的讀、寫、聽、說等能力，而最令人感到棘手的，通常是在發音的部分。同樣的一段外國話，許多人能看得懂也聽得懂，但就是無法正確流暢的唸出來，更遑論以該種外國語與他人溝通。

由於有這樣的需求，所以有些公司便推出了以矯正發音做為訴求的電腦產品。例如台灣希伯崙股份有限公司出品的 CNN 互動光碟，與法國 Auralog 公司出產的 Tell Me More。這兩種產品都可以讓外語學習者在朗讀課文時進行錄音，並顯示其波形，然後再讓學習者自行比對他們的發音波型與教學者的發音波形。

然而前述的產品卻有他們的侷限性。一方面聲音的波形對一般人並沒有特殊的意義，即使在語言方面訓練有素的專家，也無法單由觀看波形就判斷出兩個發音是否相似。另一方面，由於這些系統無法在聲音訊號中找出各個音節的所在位置，所以無法針對各個音節逐一做比對，並進而找出其中差異性較大的部分提出改善建議。這些產品在進行聲音比對的時候，只能假設教學者與學習者在同一

(請先閱讀背面之注意事項再填寫本頁)

裝 · 訂 · 線

五、發明說明(乙)

個時段內是唸到同一個音節。但是我們知道，每個人說話的速度(timing)是不同的，舉例而言，當教學者在講第5個字的時候，說不定學習者還在說第2個字，因此，以時間做為比對基礎的系統就會以教學者唸的第5個字去和學習者唸的第2個字做比較，可想而知，這樣的比對結果是不具意義的。

以下即參考第1圖來說明這樣的情形，圖1繪示的是法國 Auralog 公司出產的 Tell Me More 產品的部分使用介面。其中 標示 100 的地方顯示的是學習者要學習的外語句子。110 顯示是教學者的發音波形 120 顯示的是學習者的發音波形。雖然該產品嘗試比較教學者與學習者在唸"for"這個字上的差異(t0~t1 反白部分)，但是由於教學者與學習者在發音的速度上有所不同，所以該產品並沒有正確地找出"for"這個字在教學者發音與學習者發音中的位置。事實上，在 t0~t1 這個時段裡，教學者只唸了"for"這個字的前半部，而學習者更是沒有發出任何聲音。

之所以會有這樣的情況發生，完全是因為這類產品在比對音波時皆是採"時間(timing)"比對，是以除非學習者的說話速度皆與教學者相同，否則比對出的波形是不具意義的。

發明概述

有鑒於此，本發明提出一種自動標示音標以矯正發音的系統，包含其介面、製造方法以及使用方法。這個系統有兩個主要優點，第一，由於它能在教學者及學習者的發

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (3)

音波型上，分別標示出各個區段的音標，學習者可以更清楚的看出兩者的差異；第二，由於這個系統係依據各個區段標示之音標而知道句子中某一特定單字或音節分別出現在教學者波形及學習者波形的哪一個部分，是以可以將相對應的部分抽離出來並單獨進行比較。這些比較包含各組對應音節之間的發音差異、音高差異、強度差異、長短差異等等。

本發明的製造及使用方法可以分成三個階段—「資料庫建立階段」、「音標標示階段」、以及「發音比較階段」。在資料庫建立階段裡，我們的目標是要建立一個「音素特徵資料庫」(Phoneme Feature Database)，這個資料庫包含各個音素(語言發音的最小單位，通常對應於一個音標)的特徵資料，以做為下一階段進行標示音標時的基礎。在音標標示階段裡，我們的目標是要在一段語音波形上，標示出各個區段所對應的音標。而在發音比較階段裡，我們的目標是要對兩個已經標示出音標的波形進行比較，分析出各個對應區段間的差異程度，然後做出評分或使提出改善建議。以下我們將針對各個階段進行較詳細的說明：

在資料庫建立階段中，首先使用者必須蒐集一定數量之樣本聲音訊號，將之輸入到本系統中。這些樣本聲音訊號通常是由外語教學者所錄製的，包含許多不同文句的發音。接著，本系統將這些發音樣本切割成許多固定長度的「音訊框」(Frames)，並藉由「特徵擷取器」(Feature Extractor)分析並取得各個音訊框的各項「特徵值」(Features)。最後，

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(4)

本系統會提供一個使用介面，藉由人工判斷做分類，將屬於同一「音素」(Phoneme)的樣本音訊框蒐集在一個「音素叢集」(Phoneme Cluster)中，並自動計算每一個音素叢集中各項特徵值所共同產生之平均值與標準差，將之存入資料庫中。

在音標標示階段中，本系統所需的輸入資料是一個文句字串，以及一個由語言教學者或語言學習者針對該文句所錄製的聲音訊號。而這個階段的輸出則是一個已標示出各區段音標的聲音訊號。在做法上，本系統首先利用一個電子字典，查詢出輸入文句的對應音標，接著本系統會將輸入的聲音訊號切割成固定大小的音訊框、計算各音訊框的特徵值、並利用前一階段所得到的音素特徵資料庫，計算出每個音訊框歸屬於各個音標的機率。最後，本系統提出一個利用「動態規劃」(Dynamic Programming)方法的技術，以求得一個最佳的音標標示。

在發音比較階段中，本系統針對兩個已經在前一階段標示出音標的聲音訊號進行比對，這兩個聲音訊號通常分別來自於語言教學者與語言學習者。在做法上，我們先找出在兩個聲音訊號中相對應的部分(一個或數個音訊框)，然後將這些對應的部分逐一配對進行比較。舉例而言，如果語言學習者正在學習"This is a book"這個句子，本系統就會在教學者的聲音訊號及學習者的聲音訊號中分別找出相對於"Th"的部分進行比較，然後再找出相對於"i"的部分做比較，然後再找出相對於"s"的部分做比較，依此類推。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(ㄟ)

而比對的內容包含但不限於發音準確度、音高、強度、以及節奏。當我們比對發音準確度的時候，我們可以將學習者的發音直接與教學者比較，也可以將學習者的發音拿來與音素資料庫中該發音的資料做比較。當我們比較音高的時候，我們可以將學習者發音與教學者發音的絕對音高拿來直接做比較，也可以先計算學習者的「相對音高」(句子一部份的音高與整個句子的平均音高比)，然後再跟教學者的相對音高比較。同樣的，當我們比較發音強度的時候，我們可以將學習者發音與教學者發音在該部分的絕對發音強度拿來直接做比較，也可以先計算學習者在該部分的「相對發音強度」(句子一部份的發音強度與整個句子的平均發音強度比)，然後再跟教學者在該部分的相對發音強度比較。也同樣的，當我們比較發音節奏的時候，我們可以將學習者發音與教學者發音在該部分的時間長短直接拿來做比較，也可以先計算學習者的「相對發音長度」(句子一部份的發音長度與整個句子的總長度比)，然後再跟教學者在該部分的相對發音長度比較。

這些比較的結果，可以分別用分數或是機率百分比來表示。而經由加權計算，我們可以得出學習者整句話在發音、音高、強度、節奏上的分數，也可以更進一步，再經由加權計算出整個句子的單一分數。在進行這些加權計算的時候，各部份的分數權重可以來自於邏輯上的推斷，也可以來自於實驗所得的經驗值。

在比對及計算分數的過程中，由於本系統可以得知教

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(6)

學者與學習者在發音上的差異究竟發生在哪裡、差異的程度有多大，因此本系統也可以根據這些資訊向學習者提出改善建議。

上述系統及方法的使用介面包括：藉由音訊輸入設備而得到的聲音訊號圖，和藉由分析聲音訊號而得到強度變化圖及音高變化圖等。此外，數個區隔線段將這些圖表區隔成幾個發音區間，而每個發音區間由一個音標標註。使用者可以藉由滑鼠等輸入裝置選取一個或數個發音區間，並單獨播放那些發音區間的音訊。

在本系統中，語言學習者的聲音訊號及學習者的聲音訊號分別由一組圖表介面表示，當使用者選取教學者的聲音訊號的某些發音區間時，本系統會自動選取學習者的聲音訊號中的那些對應發音區間，反之亦然。

綜合上述，本發明是利用圖形介面比較並顯示語言學習者與語言教學者在發音上的差異，以幫助語言學習者學習正確的發音及語調。

為讓本發明之上述和其他目的、特徵、和優點能更明顯易懂，下文特舉較佳實施例，並配合所附圖示，作詳細說明如下：

圖示之簡單說明：

第 1 圖繪示的是歐洲的 Auralog 公司出產的發音練習產品之一使用介面；

第 2 圖繪示的是本發明一較佳實施例的一種自動標示音標以矯正發音之一使用者介面；

五、發明說明 (17)

第 3 圖繪示的是本發明一較佳實施例的一種自動標示音標以矯正發音之一使用者介面；

第 4 圖繪示的是本發明一較佳實施例在資料庫建立階段的系統方塊圖；

第 5 圖繪示的是本發明一較佳實施例在音標標示階段的之一系統方塊圖；

第 6 圖繪示的是本發明一較佳實施例在音標標示階段的示意流程圖；

第 7 圖繪示的是本發明在音標標示階段中進行動態比對之一示意圖；以及

第 8 圖繪示的是本發明一較佳實施例在發音比較階段的系統方塊圖。

標號說明

100：字串顯示處

110：教學者聲音訊號圖

120：學習者聲音訊號圖

200：教學內容顯示區

210：教學者使用介面

220：學習者使用介面

211，221：聲音訊號圖

212，222：音頻變化圖

213，223：強度變化圖

214，214a，214b，224：區隔線段

215：教學者指令區

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(8)

216, 226 : 音標標記區

221 : 聲音訊號圖

225 : 學習者指令區

402 : 樣本聲音訊號

404, 510 : 音訊切割器

406 : 樣本音訊框

408 : 人工音標標示器

410 : 已標示音標的樣本音訊框

412, 512 : 特徵擷取器

414 : 已標示音標的特徵值集合

416 : 叢集分析器

418, 515 : 叢集資訊

420, 514 : 音素特徵資料庫

501a : 聲音訊號

501b : 波形圖

504 : 教學內容瀏覽器

505 : 文句字串

506 : 電子音標字典

507 : 音標字串

508 : 音標標示

513 : 特徵值集合

511 : 音訊框

步驟 602 至步驟 608 係本發明之一較佳實施例之一實施步驟

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

五、發明說明(9)

較佳實施例

請參照第 2 圖，其繪示的是本發明一較佳實施例的使用者介面，其中有分 3 個部分，分別是教學內容顯示區 200、教學者使用介面 210、及學習者使用介面 220。

當使用者利用滑鼠等輸入裝置在教學內容顯示區 200 中選取一個文句字串的時候，本系統會播放對應於該文句字串且事先由教學者錄製好的聲音訊號，並在教學者使用介面 210 中顯示相關的資訊。

其中，教學者使用介面 210 包括：聲音訊號圖 211、音頻變化圖 212、強度變化圖 213、數個區隔線段 214、教學者指令區 215 及音標標記區 216。其中，聲音訊號圖 211 顯示教學者的聲音訊號的波形。強度 (intensity) 變化圖 213 是藉由分析聲音訊號的能量變化而得到的。音頻變化圖 212 是藉由分析聲音訊號的音頻 (pitch) 變化而得到的，其分析方法可以是由 Goldstein, J. S., 在 1973 年提出之 "An optimum processor theory for the central formation of the pitch of complex tones," 而得到，或是由 Duifhuis, H., Willems, L. F., 及 Sluyter, R. J., 在 1982 年提出之 "Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception," 或是 Gold, B. Morgan, N., 在 2000 年提出的 "Speech and Audio Signal Processing," 等等方法而得到。

在教學者使用介面 210 中，本系統會以區隔線段 214 將音波圖區隔成數個「發音區間」，並在音標標記區 216

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(16)

中標示各發音區間所對應的音標。舉例而言，區隔線段 214a 及 214b 間的發音區間相對於"I"的音，其音標即顯示在音標標記區 216 中該發音區間的下方。使用者可以利用滑鼠等輸入裝置選取一個或多個連續的發音區間，並經由點選教學者指令區 215 的「播放選擇部份」(Play Selected) 鈕來播放該發音區間的聲音訊號。

學習者使用介面 220 與教學者使用介面 210 類似，包括聲音訊號圖 221、音頻變化圖 222、強度變化圖 223、數個區隔線段 224、以及音標標記區 226。其功能與教學者使用介面 210 類似，如圖 3 所示，在此不再詳加贅述。但其分析的聲音訊號並非預先錄製的，而是由學習者利用學習者指令區 225 中的「錄音」"Record"鈕進行即時錄音而得到的。

如圖 3 所示，當學習者在學習者使用介面 210 中選取一段發音區間時，本系統會將該段區間以反白方式顯示，並依據標示之音標自動在教學者使用介面中選取相對應的發音區間，並同時以反白方式顯示。在這裡，我們可以看到教學者和學習者在說"great"這個單字時的時間與是不同的，但本發明仍可以分別在教學者與學習者的聲音訊號圖示上，自動而準確地標示出這個字出現的位置。

以下我們將針對此較佳實施例進行比較詳細的說明。第 4 圖繪示的是本系統在「音訊資料庫建立階段」中的主要模組。在這個階段中，「音訊切割器」404 首先將經由麥克風輸入的樣本聲音訊號 402 切割成一個一個固定長短(通

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

五、發明說明 (| |)

常是 256 或 512 個位元組)的樣本音訊框 406。緊接著，我們利用「人工音標標示器」408 以人工試聽的方式來標出每個樣本音訊框 406 的音標，至此，樣本音訊框 406 即會成為已標示出音標的音訊框 410，並將這些樣本音訊框 410 交給「特徵擷取器」412，計算出每個樣本音訊框 410 的特徵值 414。這些已標示出音標的音訊框 414 通常是一組 5 到 40 個浮點運算數，包含「倒頻譜」(Cepstrum)係數或是預測語音編碼(Linear Predictive Coding)係數等。關於音訊特徵擷取的技術可以參閱 Davis, S., and Mermelstein, P., 在 1980 年發表之 "Comparison of parametric representations of monosyllabic word recognition in continuously spoken sentences,"，或是 Gold, B. Morgan, N., 在 2000 年提出的 "Speech and Audio Signal Processing,"。

接著在「叢集分析器」416 中，我們將屬於同一音標的樣本特徵值集合 414 歸類整理成一個一個的「音素叢集」(Phoneme Cluster)，並針對每一個音素叢集，計算其特徵值集合的平均值與標準差，然後將這些叢集資料 418 存入音素特徵資料庫 420 中。關於叢集分析這方面的技術，可以參閱 Duda, R., 及 Hart, P. 所著，由 Wiley-Interscience 公司在 1973 年出版的 "Pattern Classification and Scene Analysis"。

第 5 圖所繪示的是本較佳實施例在音標標示階段中的主要模組。在這個階段中，我們的目的是要在一段聲音訊號上標示出正確的音標，然後交由教學者使用介面 210 或

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

五、發明說明 (/ 2)

學習者使用介面 220 顯示，同時也將結果交由發音比較階段中之「發音比較器」(未繪示)進行評分。這時系統需要兩項輸入資料，一個是使用者在「教學內容瀏覽器」504 中所點選的文句字串，另一個是經由麥克風輸入且對應於該文句字串之聲音訊號 501a。

由麥克風輸入的聲音訊號 501a 會經由音訊切割器 510 切割成固定大小的音訊框 511，並由特徵擷取器 512 計算出每個音訊框 511 的特徵值集合 513。音訊切割器 510 與特徵擷取器 512 的功能如前所述，在此不再重複。

在教學內容瀏覽器中選取的文句字串會經由電子音標字典 506 轉換為一個音標字串 507，舉例而言，如果使用者選取了文字字串 "This is good"，則電子音標字典會將之轉換為音標字串 "DIs Iz gud"。

我們在第 6 圖中以一個實際的例子來說明音標標示過程，當聲音訊號 501a 經由分割步驟 602 分割得到數個音訊框 511 後，會在經由特徵擷取步驟 604 進行特徵擷取而得到音訊框 511 相對應之特徵值集合，其中一個音訊框對應一個特徵值集合 513，在這些步驟進行同時，亦會對輸入之文句字串 505 進行音標字典查詢步驟 606，以得到文句字串 505 之音標字串 507，最後再由步驟 604 所擷取之特徵值集合與步驟 606 所查詢之音標字串 507 進行步驟 608 的動態比對。其中「動態比對」指的是音標標示器 508 以「動態規劃」(Dynamic Programming)法進行音標標示的工作，這個過程會將音標字串 507 中的每個音標標示到代表

(請先閱讀背面之注意事項再填寫本頁)

裝 · 訂 · 線

五、發明說明 (13)

各個音訊框 511 的特徵值集合上。這個標示過程必須符合幾個條件：第一，各個音標必須依照他們在音標字串中出現的順序逐一標示，先出現的音標先標示；第二，每個音標可能對應到零個、一個或多個特徵值集合(當一個音標對應到零個特徵值集合時，代表錄音者並未唸出那一個音)；第三，每個特徵值集合可以對應到一個音標，或是不對應到任何音標。(當一個特徵值集合不對應到任一個音標時，代表這一個特徵值集合對應於聲音訊號中的一段空白部份或是一段雜音)；第四，這個標示必須讓一個事先定義的「效用函數」(Utility Function)達到最大值(或是讓一個「懲罰函數」(Penalty Function)達到最小值)。這個效用函數所代表的是這個標示的正確程度(懲罰函數所代表的是這個標示的錯誤程度)，它可以來自於理論推斷，也可以根據實驗所得到的經驗值來推定。

第 8 圖所繪示的是以「動態規劃」(Dynamic Programming)方式進行音標標示的較佳實施例，在這裡，我們以音標字串中的各個音標做為橫軸，以聲音訊號中的各個音訊框做為縱軸，然後在表格中填入下列數值：

$\max(\text{該音訊框屬於該對應音標的機率, 該音訊框是雜音或空白的機率})$

其中各音訊框屬於各個音標或是雜音及空白的機率，可以藉由參照音素資料庫而得到。基本上，我們將各個音

(請先閱讀背面之注意事項再填寫本頁)

裝訂線

五、發明說明(14)

訊框的特徵值集合與音素資料庫中各個音素(一個音標對應於一個音素)的特徵值集合的平均數與標準差做比較，經由簡單的數學運算即可得到這些機率。關於這方面的技術，可以參閱 Duda, R., 及 Hart, P. 所著，由 Wiley-Interscience 公司在 1973 年出版的 "Pattern Classification and Scene Analysis"。

此外，如果在某儲存格的資料是來自於該音訊框是雜音或空白的機率時，我們會在該儲存格加上特別的標記。在第 7 圖中，我們是以灰階網底來標示這些儲存格。

接下來我們必須在第 7 圖的動態比對表中找到一條由左上角至右下角的路徑，這條路徑所代表的就是音標標示的結果。舉例而言，在第 7 圖中第一個音標 ϕ 對應於音訊框 1 與 2，第二個音標 l 對應於音訊框 3 與 4，而第三個音標 s 則對應於音訊框 5 與 6。

這條路徑必須符合幾個條件：第一，這條路徑只能往右、往右下、或往下行進。第二，這條路徑所代表的音標標示必須能讓我們所定義的效能函數達到最大值，也就是說，這個路徑必須代表一個最佳的音標標示。

如果這條路徑經過一個以灰階標示的音訊框，則代表這個音訊框是一個雜音或是空白訊號。否則，當這條路徑往右行進時，代表接下來音標並未在這個聲音訊號中出現；當這條路徑往右下行進時，代表前後兩個相鄰的音訊框剛好對應於兩個相鄰的音標；而當這條路徑往下行進時，則代表前後兩個音訊框對應於同一個音標。

(請先閱讀背面之注意事項再填寫本頁)

裝訂線

五、發明說明(15)

在這裡，我們可以將效能函數定義成這條路徑在動態比對表中，在往下及往右下行進時所經過的各個機率值的乘積(當這個路徑往右行進時，代表我們將略過那一個音標，因此代表那一個音標的機率值不應該計入我們的效能函數中)。理論上，這個乘積相當於這條路徑是正確的音標標示的機率。

這樣的一條路徑，可以利用動態規劃法(Dynamic Programming)得到，關於以動態規劃法解決這類問題的技術，可以參考 J. Ullman 於 1977 年在 Computer Journal 10, pp141-147 所發表的 “A Binary n-gram technique for automatic correction of substitution, deletion, insertion, and reversal errors in words.” 或是 R. Wagner 與 M. Fisher 於 1974 年在 Journal of ACM 21, pp168-178 所發表的 “The String to String Correction Problem.”

第 8 圖所繪示的是本系統在發音比對階段中的主要模組。在這個階段中，本系統先就發音、音高、強度、節奏等四個部份分別進行評分，並列出改善建議。接著，我們再以加權的方式從這四個分數算出一個總分。至於加權的比重，可以來自於理論推斷，也可以來自於實際經驗。

如前所述，在這些評分的過程中，本系統會先找出在兩個聲音訊號中相對應的部分(一個或數個音訊框)，然後將這些對應的部分逐一配對進行比較。舉例而言，如果語言學習者正在學習 “This is a book” 這個句子，本系統就會在教學者的聲音訊號及學習者的聲音訊號中分別找出相對

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(16)

於”Th”的部分進行比較，然後再找出相對於”i”的部分做比較，然後再找出相對於”s”的部分做比較，依此類推。而如果一个音標(或音節)在一个聲音訊號中對應於多個音訊框，我們可以先求得這些音訊框在特徵值(用來比較發音)、音高、強度、以及長度上的平均值，然後再與另一個聲音訊號中相對求得的平均值做比較。我們也可以將來自於教學者與來自於學習者的各個音訊框逐一配對做比較，以分析在同一音標範圍內，發音、音高、以及強度隨著時間所顯現的變化。

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

六、申請專利範圍

1.一種自動標示音標以矯正發音之方法，包括：

一音素特徵資料庫建立步驟，包括利用樣本聲音訊號建立複數個音素叢集，其中一個音素叢集對應一個音標；
一音標標示步驟，包括：

分割一聲音訊號成複數個音訊框，並計算出每一個音訊框的特徵值集合；以及

依據每一個音訊框的特徵值集合，判斷該音訊框之所屬音素，並予以標示相對的音標；以及

一發音比較步驟，包括比較兩個聲音訊號中相對於同一音標的各組音訊框，做出評分並提出改善建議。

2.如申請專利範圍第1項所述之自動標示音標以矯正發音之方法，其中的音素資料庫中包含複數個音素叢集，而每一個音素叢集對應於一個音標，而該音素叢集的資料是藉由分析對應於該音素的樣本音訊框而得到。

3.如申請專利範圍第2項所述之自動標示音標以矯正發音之方法，其中音素資料庫之建立方法包括：

輸入樣本聲音訊號；

分割樣本聲音訊號成複數個樣本音訊框；

由人工試聽的方式判斷各音訊框所屬的因素叢集，並標示相對於該音素的音標；

分別計算各個樣本音訊框的特徵值集合；以及

對於每個音素叢集，計算其所屬的樣本音訊框的特徵值集合的平均值及標準差。

4.如申請專利範圍第2項所述之自動標示音標以矯正發

六、申請專利範圍

音之方法，其中輸入的樣本聲音訊號經音訊切割器分割成數個音訊框之後，由人工試聽的方式判斷各音訊框所屬的音素叢集，並標示相對於該音素的音標。

5.如申請專利範圍第2項所述之自動標示音標以矯正發音之方法，其中每一音素叢集的資料包含所有對應於該音素的音訊框的特徵值集合的平均值及標準差。

6.如申請專利範圍第 1 項所述之自動標示音標以矯正發音之方法，其中的音標標示步驟包括：

輸入一文句字串及對應於該文句字串之一聲音訊號；
藉由一電子音標字典，查得輸入文句字串所對應的複數個音標；

分割該輸入聲音訊號成複數個音訊框；

分別計算各個音訊框的特徵值集合；

依據一音素特徵資料庫所包含之複數個音素叢集資訊，計算各個音訊框屬於輸入文句字串所對應的各個音標的機率；

根據各音訊框屬於各個音標的機率，求得一最佳音標標示，該音標標示是所有可能的音標標示中，最有可能是正確的音標標示者；以及

顯示各音訊框所對應之音標。

7.如申請專利範圍第6項所述之自動標示音標以矯正發音之方法，其中音標標示係藉由比較輸入字串及其相對應的輸入聲音訊號而得到。

8.如申請專利範圍第6項所述之自動標示音標以矯正發

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

六、申請專利範圍

音之方法，其中即使在輸入字串所對應的某些音標並未出現在輸入的聲音訊號中的狀況下，仍能正常工作，並標示出其他出現的音標。

9.如申請專利範圍第6項所述之自動標示音標以矯正發音之方法，其中即使在輸入的聲音訊號中的某些區段是多餘而不對應於輸入字串的任何部分的狀況下，仍能正常工作，並標示出該輸入聲音訊號其他部分的音標。

10.如申請專利範圍第 6 項所述之自動標示音標以矯正發音之方法，其中求得最佳音標標示的方法係採用一動態規劃法技術。

11.如申請專利範圍第 10 項所述之自動標示音標以矯正發音之方法，其中該動態規劃法技術係使用一比較表，該比較表的縱軸(或橫軸)為輸入字串所對應的各個音標，而橫軸(或縱軸)則是經切割輸入聲音訊號所得的各個音訊框，或對應於各各音訊框的特徵值集合。

12.如申請專利範圍第 11 項所述之自動標示音標以矯正發音之方法，其中最佳音標標示的求得方法，是在比較表中尋找一條由左上至右下(或由右下至左上)的路徑，而該路徑使得一個事先定義好的效能函數達到最大值(或是讓一個「懲罰函數」達到最小值)。

13.如申請專利範圍第 1 項所述之自動標示音標以矯正發音之方法，其中發音比較步驟所比較的兩個聲音訊號，其一為預先錄製的聲音訊號，其一為即時錄製的聲音訊號。

六、申請專利範圍

14.如申請專利範圍第 1 項所述之自動標示音標以矯正發音之方法，其中發音比較階段所比較的项目包括發音準確度、音高、強度及節奏等之比較。

15.一種自動標示音標以矯正發音之使用者介面，包括：

一聲音訊號圖，係藉由一音訊輸入設備而得到；

一強度變化圖，係分析該聲音訊號圖而得到；

一音頻變化圖，係分析該聲音訊號圖而得到；

複數個區隔線段，其中兩兩該些區隔線段形成一發音區間，一個發音區間對應一個音標之發音時間；以及

一音標標記區，係顯示該些發音區間所對應之該些音標；

其中可標記至少一發音區間，以命令發出該發音區間之音標聲音。

16. 如申請專利範圍第 15 項所述之自動標示音標以矯正發音之使用者介面，包括顯示該聲音訊號之一音頻變化圖及一強度變化圖。

17.如申請專利範圍第 15 項所述之自動標示音標以矯正發音之使用者介面，包括結合相鄰且歸屬同一該音素叢集的複數個音訊框，以區隔成同一發音區間。使用者可選取一個或多個發音區間，並要求系統播放相對於該發音區間的聲音訊號。

18.如申請專利範圍第 15 項所述之自動標示音標以矯正發音之使用者介面，當使用者在一個聲音訊號圖上選取

(請先閱讀背面之注意事項再填寫本頁)

裝
訂
線

六、申請專利範圍

一個或多個連續的發音區間時，系統會自動在另一個聲音訊號圖上選取相對應的發音區間。

19.如申請專利範圍第 15 項所述之自動標示音標以矯正發音之使用者介面，其中該聲音訊號圖係以音訊框為最小選取及處理單位。

20.一種自動標示音標以矯正發音之系統，包括：

一輸入設備，係輸入一文句字串及對應於該文句字串之一聲音訊號；

一電子音標字典，用以查閱得到對應於文句字串的音標字串；

一音訊切割器，係分割該聲音訊號成複數個音訊框；

一特徵擷取器，連接該音訊切割器，係從該些音訊框擷取相對應之特徵值集合；

一音素特徵資料庫，包括複數個音素叢集，其中一個音素叢集對應一個音標；

一音標標示器，連接該特徵擷取器、該電子音標字典及該音素特徵資料庫，係依據音素特徵資料庫內含之複數個音素叢集，計算該些音訊框為該文句字串之該些音標之複數個可能機率，將該些音訊框之該些可能機率標示在一動態比對表中，以及依據該動態比對表之一動線方向確定該些音訊框對應之該些音標；以及

一輸出設備，顯示輸入聲音訊號的波形圖、音頻變化圖、強度變化圖、以及對應於各個發音區間的音標等。

Shopping Mall

Pronunciation Exercise

Listen

Speak

Wait

Thanks

for

your help.

100

110

120

10

10

10

your pronunciation

10

11

12

13

No.

Score

10


11

12

13

第1圖

Teacher Profile



Teachers
Stephanie Hwang
Judy Wu
James Lin
Katherine Lin

o Name: Yelling Lin


o Birth: Kichang, Taiwan

o Age: 36

o Sex: Male

財國法入語言訓練與測驗中心

The Language Training & Testing Center



口說能力測驗










My test number is _____ and my seat number is _____

I am _____

How are Cindy and the kids? Have they started back to school yet? You know, we haven't.

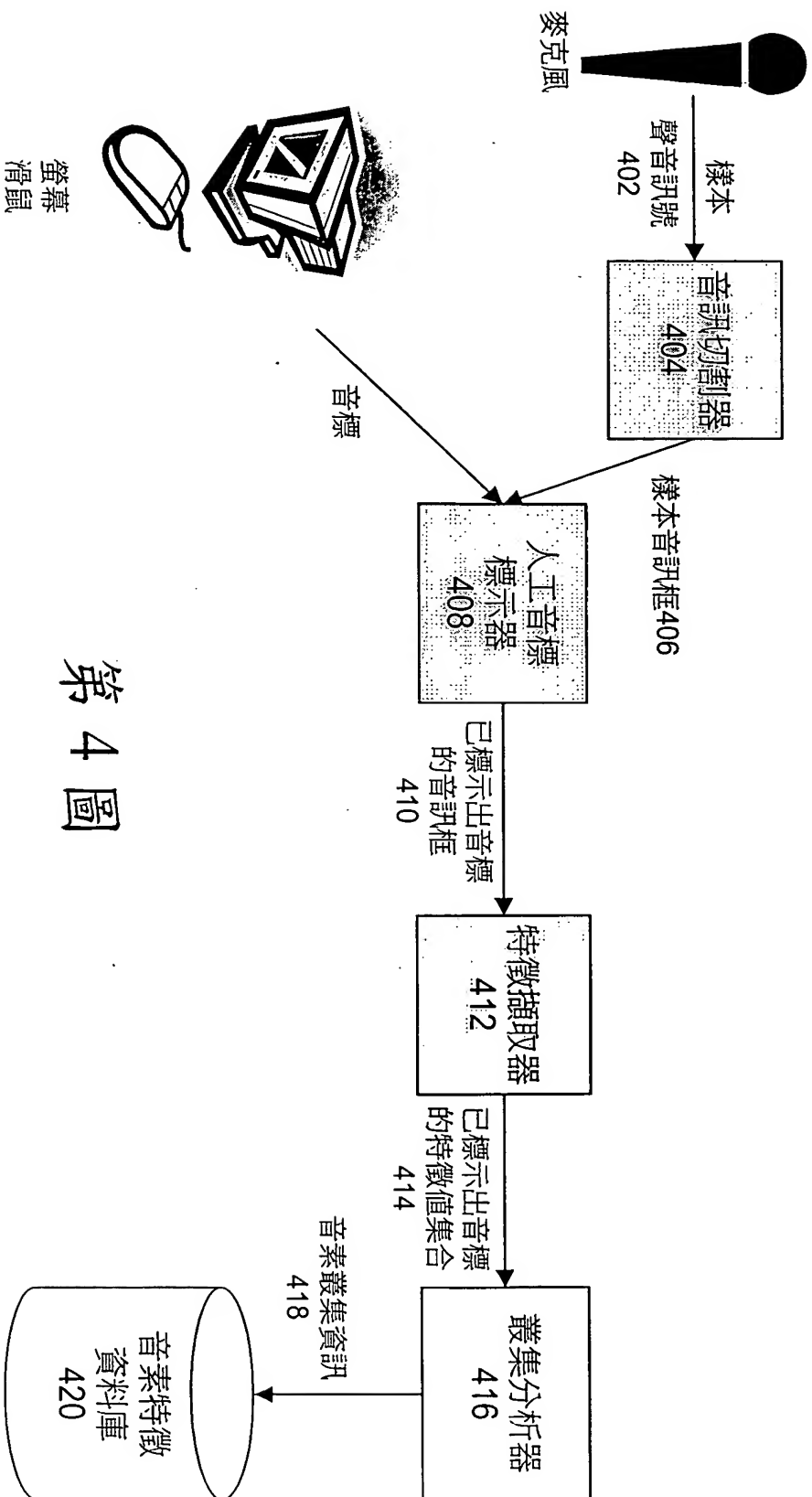
It's great to see you

It's so good to see you

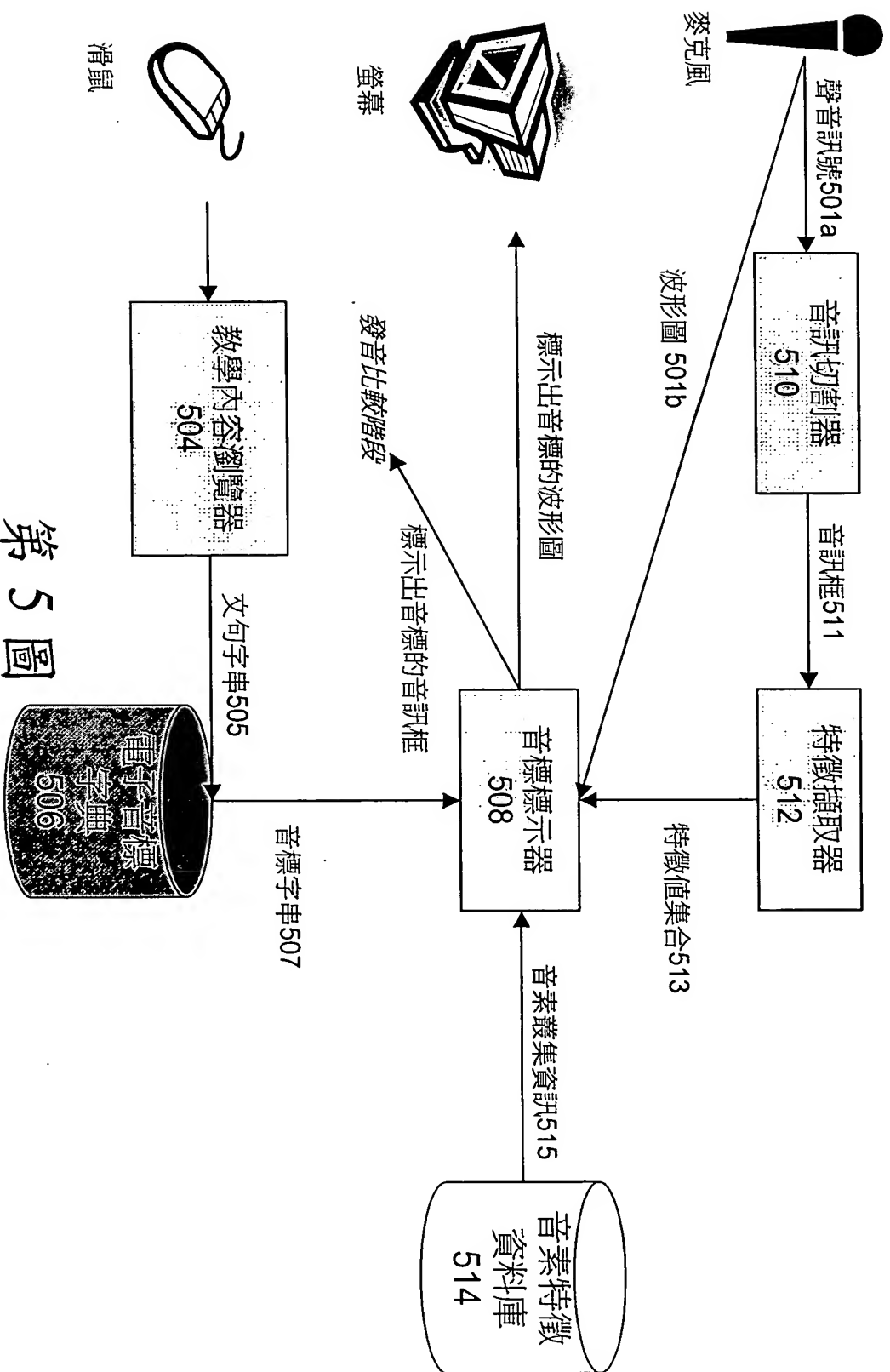


224 221 223 220 226

第 3 圖



第4圖



第 5 圖

聲音訊號
501a



分割步驟602

音訊框
511

特徵擷取步驟604

$\begin{bmatrix} 3.2156 \\ 4.1287 \\ \dots \\ 19.654 \end{bmatrix}$	$\begin{bmatrix} 3.2156 \\ 4.1287 \\ \dots \\ 19.654 \end{bmatrix}$	$\begin{bmatrix} 3.2156 \\ 4.1287 \\ \dots \\ 19.654 \end{bmatrix}$	$\begin{bmatrix} 3.2156 \\ 4.1287 \\ \dots \\ 19.654 \end{bmatrix}$	$\begin{bmatrix} 3.2156 \\ 4.1287 \\ \dots \\ 19.654 \end{bmatrix}$
---	---	---	---	---

特徵集



動態比對步驟608

音標字串
507

ðɪs ɪz ðə ɜθ

文字字串
505

This is the earth

音標字典查詢步驟606

第 6 圖

	ǝ	ɪ	s	ɪ	z	ǝ	ə	ʒ	θ
音訊框1	0.5	0.04	0.1	0.04	0.2	0.5	0.04	0.04	0.2
音訊框2	0.6	0.03	0.03	0.03	0.3	0.6	0.03	0.03	0.1
音訊框3	0.2	0.4	0.1	0.4	0.2	0.2	0.02	0.02	0.1
音訊框4	0.05	0.8	0.05	0.05	0.05	0.05	0.1	0.1	0.05
音訊框5	0.1	0.1	0.7	0.1	0.2	0.0	0.0	0.0	0.0
音訊框6	0.1	0.01	0.7	0.01	0.1	0.01	0.01	0.01	0.01
音訊框7	0.1	0.03	0.2	0.6	0.03	0.1	0.03	0.1	0.03
音訊框8							
音訊框9	...								
音訊框N	0.6	0.03	0.03	0.03	0.3	0.6	0.03	0.03	0.1

第7圖